Chapter
10

Matching Census Database and Manitoba Health Care Files

Christian Houle, Jean-Marie Berthelot, Pierre David, and Michael C. Wolfson, Statistics Canada; Cam Mustard and Leslie Roos, University of Manitoba

Abstract

Introduction*: In the current economic context, all partners in health care delivery systems, be they public or private, are obliged to identify the factors that influence the utilization of health care services. To improve our understanding of the mechanisms that underlie these relationships, Statistics Canada and the Manitoba Centre for Health Policy and Evaluation have set up a new database. For a representative sample of the population of the province of Manitoba, cross-sectional microdata on individuals' health and socio-economic characteristics were linked with detailed longitudinal data on utilization of health care services.*

Data and methods*: The 1986-87 Health and Activity Limitation Survey, the 1986 Census and the files of Manitoba Health were matched (without using names or addresses) utilizing a CANLINK software. In the pilot project, 20,000 units were selected from the Census according to modern sampling techniques. Before the files were matched, consultations were held and an agreement signed by all parties to establish a framework for protecting privacy and preserving the confidentiality of the data.*

Results*: A match rate of 74% was obtained for private households. A quality evaluation based on the comparisons of names and addresses over a small subsample established that the overall concordance rate among matched pairs was 95.5%. The match rates and concordance rates varied according by age and household composition. Estimates produced from the sample accurately reflected the socio-demographic profile, mortality, hospitalization rate, health care costs, and consumption of health care by Manitoba residents.*

Discussion*: The match rate of 74% was satisfactory in comparison with response rates reported by the majority of population surveys. Because of the excellent concordance rate and the accuracy of the estimates obtained from the sample, this database will provide an adequate basis for studying the association between socio-demographic characteristics, health and health care utilization in province of Manitoba.*

Introduction

A number of studies have clearly shown that there is a link between an individual's socio-economic status and the probability of his or her death during a given period of time (Wolfson et al., 1993; Marmot, 1986; Wilkins et al., 1991). Other studies have shown that the prevalence of certain diseases varies greatly depending on the socio-economic characteristics of the area in which an individual resides (Anderson, 1993, Dougherty, 1990). In addition, several Canadian surveys have already provided cross-sectional data on individuals' health status and socio-economic status, along with self-reported information

on the use of health services, e.g., General Social Survey of 1991 (Statistics Canada, 1994a), Ontario Health Survey of 1990, Enquête Santé Québec of 1987 and 1992-93, Health and Activity Limitation Survey of 1986 and 1991 (Statistics Canada, 1988), Canadian Health and Disability Survey of 1983-84 (Statistics Canada, 1986a), Canada Health Survey of 1978-79 (Health and Welfare Canada, 1981). However, to our knowledge, there is no Canadian longitudinal database that combines information on health, use of health services, and socio-economic characteristics. In an effort to meet this information need, Statistics Canada and the Manitoba Centre for Health Policy and Evaluation (MCHPE) set up a joint pilot project to evaluate the possibility of creating such a database using existing data sources.

The primary objective of the pilot project was to evaluate the feasibility of combining the following three data sources: the 1986 Census of Population, the 1986-1987 Health and Activity Limitation Survey (HALS) , and the Manitoba Health (MH) longitudinal file on health care service utilization. The database resulting from this combination will enable researchers to explore new directions with respect to health determinants. In this article, we describe the matching of files, the selection of the sample for analysis purposes and the results, which show the representativeness of the database created and validate the techniques employed.

Confidentiality and Right to Privacy

When creating a database from both administrative and survey data, it is essential to ensure the confidentiality of the data and prevent any unwarranted intrusion into individuals' privacy. In accordance with the policies of the collaborating agencies, certain procedures were undertaken prior to matching these data sets. They include consultations with the Privacy Commissioner of Canada, the Faculty Committee on the Use of Human Subjects in Research at the University of Manitoba, and Statistics Canada's Confidentiality and Legislation Committee. In addition, Manitoba Health's Committee on Access and Confidentiality was informed of the project.

Following these consultations, and in accordance with the formal policies of Statistics Canada, the Minister responsible for Statistics Canada authorized the matching as outlined below:

- A pilot project for evaluating the feasibility and utility of data matching.
- It was explicitly stated that individuals' names and addresses would not be used for matching purposes, nor would they appear in the database.
- The matching would be done entirely on the premises of Statistics Canada by persons sworn in under the Statistics Act.
- Only a sample of 20,000 matched units would be used for purposes of research and analysis; and
- Access to the final data would be strictly controlled in accordance with the provisions of the Statistics Act. In addition, all activities with the linked data set are covered by a memorandum of understanding including Statistics Canada, the University of Manitoba and the Manitoba Ministry of Health.

Data

The detailed questionnaire (questionnaire 2B) of the 1986 Census of Population contains extensive socio-economic information including variables such as family composition, dwelling characteristics, tenure, ethnic origin and mother tongue, as well as a number of variables relating to income and educational attainment (Statistics Canada, 1986b). This questionnaire was filled by persons residing in Manitoba on

June 3, 1986 in a proportion of approximately one household in five. The other households completed a short form designed solely for enumerating the population. Thus, the file used for matching purposes consisted of 261,861 records. The individuals represented by these records lived in two types of dwellings: private or collective. While this article focuses primarily on the private household component, there were in 1986 more than 26,161 persons in Manitoba living in a collective dwelling according to the Census. Examples of collective dwellings are hospitals, hospices, nursing homes, institutions for the physically handicapped, orphanages, psychiatric institutions, hotels/motels, work camps, jails, Hutterite colonies, military residences, religious institutions, student residences and YMCAs.

The 1986 HALS was a postcensal survey that sought to identify individuals who, because of their health, were limited by the type or amount of daily activities that they could perform. A postcensal survey refers to a question from the Census (in this case, Question 20 on disabilities) which serves to enrich the survey sample by identifying a high proportion of the target population. An appropriate questionnaire was then completed for each person sampled. For HALS, the Manitoba population living in private households and having disabilities was studied on the basis of a sample of 5,480 persons representing a population of 150,857 persons having at least one disability. The data set created, contained information on individuals' health and functional limitations as well as on type of employment, educational level, transportation, housing and recreation. Since the survey was of the self-reporting type, the data represent the situation of respondents from their viewpoint rather than from an administrative or clinical viewpoint.

The MH longitudinal file, for its part, contains information on visits to physicians, stays in hospital, diagnoses, surgical procedures, admission to personal care (nursing) home, health care received at home, the date and cause of death, and other data on health care utilization. A number of innovative studies in health care research have used this file (Roos et al., 1987, Shapiro and Roos, 1984). For this pilot project, a register of persons covered by Manitoba health insurance was identified from June 1986, using the date of commencement of health insurance coverage and the date of cancellation of coverage. The register contained 1,047,443 records.

Methods

The matching project was divided into three main stages. The first stage consisted of pairing individuals belonging to three distinct data sources. The second stage consisted of assessing what proportions of the pairs formed represented the same individual. The third stage consisted of selecting a sample of 20,000 matched units used to create the database for analysis purposes. In this section, we shall deal with the methodology used in each of these stages in turn.

Matching

The CANLINK system (Smith, 1981; Fellegi and Sunter, 1969) developed at Statistics Canada, was used for the pairing stage. CANLINK is a probabilistic matching software that pairs records from two sets of data by using the discriminatory power of the common variables available. The software weights the pairs of records according to the degree of concordance of the values observed and also takes account of the probability of random concordances. The files paired were that of the 2B sample from the 1986 Census covering the province of Manitoba and the file of persons registered with MH in June 1986, containing a subset of the variables available. Only these two files were involved in the probabilistic matching, since the 1986-1987 HALS sample was drawn from the Census 2B sample (Dolson et al., 1987), and all HALS records were already paired to those in the Census by a unique identifier.

The individual records which were paired came from two files, one containing the records of 261,861 individuals living in Manitoba (derived from the 2B file of the 1986 Census), and the other, containing the records of 1,047,443 persons (a derivative of the Manitoba Health file). The strategy adopted for identifying

pairs representing the same individual (good pairs) consisted of dividing up the two data sets into blocks and forming only pairs of individuals belonging to the same block.

The pairs of records were compared only if all the blocking variables concurred. It was therefore necessary to choose carefully so as not to eliminate at an early stage a great number of "good" pairs. It will be recalled that the most discriminant variables, namely surnames, given names and addresses, were not used in this study. Because of this constraint, we were forced to choose other combinations of variables that were limited in discriminatory power and then apply innovative techniques.

Two matching phases were carried out. First, after examining various possible definitions, we defined a block as a set of four individual characteristics, namely a person's sex, year of birth, month of birth and postal code. In the second matching phase, the definition was relaxed in order to form more pairs of individuals. The exact year and month of birth were replaced by the person's age, which made it possible to compare an individual with a greater number of candidates. In addition, the area covered by the geographic variable in urban settings was expanded by a factor of approximately three, with the postal code being replaced by the census enumeration area.

Through these matchings the census file was divided into three subsets: records which had clearly matched (definites), those which had matched but for which the discriminatory power of the available variables raised a doubt (based on CANLINK criteria (possibles)), and those which had never matched.

While the information on family structure was used in the matching process, the CANLINK system compared only two individuals at a time, without taking account of matches obtained for other family members. We had to define a series of rules in order to ensure the consistency of matchings within a given family and between two matching phases (David et al., 1993).

Evaluation of the Concordance of the Pairs Formed

The evaluation pursued two objectives. First, it was important to determine the degree of accuracy with which we had associated the Manitoba Health data with the Census data (definite matches only). Then it was necessary to assess whether the rules that had been developed for rejecting certain "possible" matches were adequate.

A sample of 1,000 families was drawn, representing 2,102 matched individuals. As stratification variables, we used urban/rural area as determined by Census, family composition (person living alone, couple with child, couple without child, multiple family) and matching status (definite or possible). MH extracted the names and addresses of all these individuals and their family members. It should be understood that this identifying information was not used to determine the validity of specific matches, but only to estimate actual matching rates at aggregated levels. Names and addresses were compared manually with those on the mircrofilmed 2B questionnaires kept at Statistics Canada.

Sample Selection

As the project involved three databases, the sampling frame was derived from the 2B file from the 1986 Census. The sample size was already set at a maximum of 20,000 units and the database created had to combine information from the census files and the MH file, as well as information from the HALS file. The HALS sample used the individual as its sampling unit, whereas the analysis of the overall population of Manitoba used the household as defined by the Census. Several options were considered in order to try to construct a single database, however the complexity of the analysis would have negated any potential gains in accuracy. To ensure a balance between simplicity of analysis and an effective design, the selection process consisted of constructing two independent databases: the first, to study the link between disability, socio-economic status and health, and the second, to analyse the general population of Manitoba. To

maximize the use of the 20,000 units, it was also necessary to take account of the overlap between these two databases.

Owing to the complex sample design of HALS, the relatively small number of individuals sampled and the importance of this database from an analytical standpoint, matched individuals with disabilities were all selected. These accounted for 4,434 basic units. This sample formed the first database, used for the analysis of persons with at least one functional disability.

There were therefore 15,566 units left to form the general population database plus the expected number of units overlapping the two databases. Still pursuing the objective of optimizing the sample design, an evaluation indicated that stratification was appropriate. Stratification has several advantages. First, it serves to reduce the overall variance of the estimates. Second, it ensures a standard of quality for estimates relating to subgroups of interest in the population. Third, stratification can result in improved accuracy for cases in which non-sampling error can be taken into account. Finally, stratification is especially effective when the stratification variables are correlated with the target variable.

Since a number of studies have established links between socio-economic status and health, it was natural to use socio-economic variables to construct the strata. In addition, there was no disadvantage to using the household as the sampling unit, since socio-economic status is generally the same for all members of a given household. Since it was the 1986 Census file that entirely determines the composition of this population, all the stratification variables were either taken directly from that file or derived from it. The final number of strata for private households was 611. The total number of units drawn was 16,387. These represented 46,670 persons.

Finally, it is common practice to adjust sampling weights so that the totals estimated by the sample will reflect as accurately as possible the counts of the population studied. With post-stratification, the counts can be adjusted for categories for which the number of units was insufficient to create a real stratum but which were of sufficient analytical importance to justify the use of special techniques. These techniques changed the initial weight subject to the constraints of minimum change (Kovacevic, 1995). For private households, the counts by age group, rural or urban geographic area, marital status and sex were used to adjust the weights to the individual level, while rural or urban geographic area, household size and tenure played the same role for adjusting at the household level.

Results

Results for Matching

Despite the conservative approach applied to the initial matches, overall, 74% (174,476 out of 235,700) of individuals from the census living in a private household were matched with an individual in the Manitoba file. This rate varied according to geographic mobility, age, marital status and family size.

The factors that had the greatest influence on the match rate were all related to individuals' geographic mobility. Hence, the following groups of individuals were more difficult to match: young adults (between 20 and 25 years of age: Figure 1), persons who had changed their place of residence between the 1981 and 1986 censuses (Table 1), and divorced or separated persons (Table 2). Among these groups, frequent changes of address and family structure made concordance between the two data sources more difficult than among less mobile groups. The reason for this is that since the Census figures date from June 3, 1986, and some MH data are dated December 31, 1986, there was more likely to be an information lag with respect to mobile individuals.

Figure 1. -- Match Rate by Age
Private Households Only

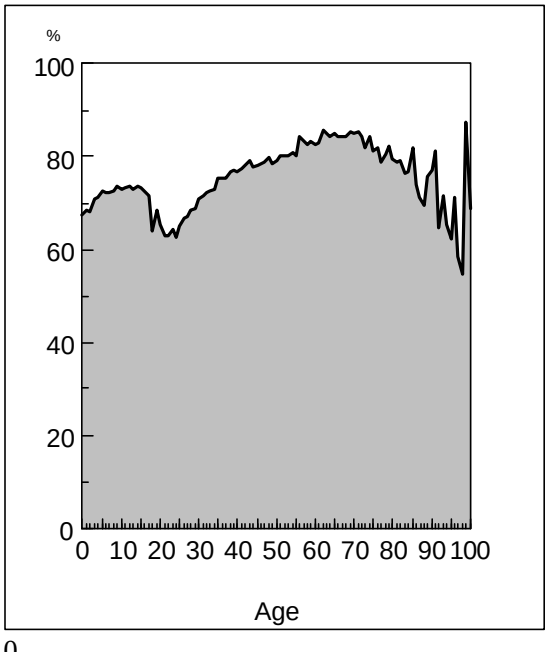

Table 1. -- Match Rate According to Mobility:
Private Households Only

Mobility	Match Rate %
Same household	81.7
Same CD	65.8
Other CD	62.5

CD: Census Division, a geographic unit used by the Census. Manitoba is made up of twenty-three Census Divisions.

Table 2. -- Match Rate According to Marital Status:
Private Households Only

Marital Status	Match Rate %
Married	78.5
Widowed	74.5
Single	71.2
Divorced	61.4
Separated	43.4

The effect of age on the match rate was not surprising. Children under fifteen years of age and adults between thirty and sixty years of age had better rates, owing to their more stable situation. Among individuals over 85 years of age, there was greater variability in the data due to the rate of institutionalization and the small number of cases.

Among individuals who did not move between the 1981 and 1986 censuses (same household), one might have expected an even better match rate. The rate of 81.7% is perhaps an indication that using the methodology described thus far, there is an upper limit of around 80% on matches, given that the files are not totally free of errors.

Intuitively, family size is correlated in two opposite ways with the match rate. While a large family has an intrinsic constraint on the mobility of the family nucleus, some members of the family will periodically attach or re-attach themselves to this nucleus. Table 3 indicates that match rate dropped off significantly as family size increased.

Table 3. -- Match Rate According to Family Size:
Private Households Only

Family Size	1	2	3	4	5	6	7	8	9	10+
Match rate %	66.4	78.5	74.7	79.8	77.1	70.0	55.9	51.4	39.2	46.7

Results for Evaluation of Concordance of Pairs Formed

Table 4 shows that overall, more than 95% of the definite matches retained represented the same individual. As the sample of 20,000 units was drawn from definite matches only, this meant that the matching was of exceptional quality. Also, due to the fact that the rate of concordance of names among possible pairs was only 40% indicated that the enforcement of a conservative methodology was justified and prudent, as they prevented a large proportion of bad matches.

Household size was closely related to the concordance rate. Persons living alone and those living in households of eight or more persons exhibited a lower concordance rate, namely 86.8% and 90.8% respectively. These results would seem to be due to the small number of discriminant variables available for persons living alone and the fact that in the case of large households, there was often more than one family within the household.

Table 4. -- Rate of Concordance of Names According to Various Groupings:

Private Households Only

Match Status	Concordance on Names %	Standard Error* %
"Possible" match"	40.1	2.4
"Definite" match"	95.5	0.5
Indian reserves	95.3	1.5
Household of 1 person	86.8	2.5
Household of 2 to 7 persons	96.5	0.5
Household of 8 or more persons	90.8	2.0

** The design effect is ignored in calculating the standard error.*

A final point to be observed is that ***matched*** inhabitants of Indian reserves [1] had a concordance rate equivalent to that of persons living off reserve.

Results for Sample Selection

Overall, we were pursuing two specific objectives in designing the stratification. First, it was necessary to come as close as possible to having a self-weighted design so as to allow for the use of existing computer applications. The costs of custom applications and the time required to develop them would have been a major handicap for any subsequent analysis. The first objective was attained by avoiding oversampling of strata to the extent possible and by maintaining a certain uniformity of weights within each stratum formed. The second objective was to use socio-economic variables in the process of forming strata. Therefore, when stratum sizes permitted, we used variables derived from income, education level, family structure, age and geography.

There were major conceptual differences with respect to the definition of the populations represented by the Census and by the MH file. Only persons having a usual place of residence in Manitoba on June 3, 1986 were enumerated in that province. The MH file was made up of all persons who were covered by the health insurance plan. Some of these persons no longer lived in Manitoba or may not have indicated a change in their status, resulting in some overcoverage. The MH file contained no information on residents of several categories of collective dwellings for which medical services were provided by the federal government, such as military camps and some Indian reserves, whereas the Census considered these persons to be residents of Manitoba.

For purposes of comparison, we excluded persons living in nursing homes (an institutional collective dwelling) from the MH counts in the tables that follow. According to the census definition, persons who had stayed for 180 days or more in a health care institution were considered institutionalized, and therefore excluded. Despite efforts to make the two universes uniform, the fact remains that we managed only to approximate the counts of persons living in institutions. Consequently, the populations compared represent ***approximately*** those persons living in a private household or a ***non***-institutional collective household.

As Table 5 shows, despite major conceptual differences, the sizes of the two populations by age group were quite comparable. Overall, the estimated total sizes of the two populations differed by only 0.1%, although males were underestimated by 1.2% and females overestimated by 1.4%. It was also observed that the greatest differences were amongst younger individuals.

Table 5. -- Accuracy of the Sample by Age Group Versus MH:
Private And Non-Institutional Collective Households

Age	Males MH	Difference from Sample %	Females MH	Difference from Sample %	Total MH	Difference from Sample %
0 to 4 years	32 743	2.57	31 105	1.98	63 848	2.28
5 to 14 years	78 076	1.47	73 912	2.87	151 988	2.15
15 to 24 years	86 722	-1.24	82 971	1.61	169 693	0.15
25 to 44 years	165 783	-2.84	159 458	1.92	325 241	-0.51
45 to 64 years	96 989	-1.71	98 997	0.92	195 986	-0.38
65 years and +	57 904	-1.34	74 129	-1.10	132 033	-1.20
Total	518 217	-1.20	520 572	1.39	1 038 789	0.10

Tables 6, 7 and 8 compare the mortality rate, medical care utilization and hospital care utilization by whether they were estimated from our sample or from the MH file. It should be noted that the death rates reported in the literature (Statistics Canada, 1994b) were slightly higher than those presented in Table 6, with the difference increasing with age. This may be explained by the fact that our files exclude individuals living an institutional collective dwellings, who exhibit a higher mortality rate than persons living in private households.

Table 6. -- Annualized Mortality Rates Based on the Period from June 1986 to May 1989:
Private and Non-Institutional Collective Households

Age	Annual Mortality Rate* (x 1,000) MH	Annual Mortality Rate* (x 1,000) Sample	95% Confidence Interval for the Sample
0-4	0.51	0.02	(0, 0.18)
5-44	2.49	2.04	(1.54, 2.54)
45-49	3.23	2.78	(0.57, 4.99)
50-54	5.12	3.68	(1.03, 6.33)
55-59	8.42	8.91	(4.80, 13.02)
60-64	12.43	10.48	(6.01, 14.95)
65-69	18.96	19.03	(12.69, 25.37)
70-74	28.63	24.59	(16.76, 32.42)
75-79	42.77	43.09	(30.85, 55.33)
80 and over	76.98	73.67	(57.41, 89.93)
Total	6.71	6.11	(5.39, 6.83)

**Number of estimated deaths (over the three years period) divided by estimated population total times 3.*

Overall, the mortality rate estimated by the matched sample (6.11) was lower than the one derived from the MH file (6.71), but this difference was not statistically significant at a 95% confidence level. It should be noted that the confidence intervals derived from our sample contained the value calculated by MH for all age groups except children aged 0 to 4. While the number of deaths in this category were relatively small, the difficulty in matching children under one year of age may be related to this underestimate. Additionally, this also indicates that any analysis specific to children from 0 to 4 years of age be conducted with caution, especially where the prevalence of a disease or condition was low.

Table 7. -- Number and Costs of Medical Services, 1986-87 Fiscal Year:
Private and Non-Institutional Collective Households

Selected Type of Practice	Number of Services			Costs of Services ($)		
	MH	Sample	Relative Difference (%)	MH	Sample	Relative Difference (%)
Internal medicine	699,542	702,735	0.46	19,060,658	18,904,922	-0.82
Paediatrics	416,157	449,122	7.92	6,932,217	7,359,290	6.16
Psychiatry	154,279	146,704	-4.91	8,970,489	8,468,584	-5.60
Surgery	406,907	409,097	0.54	21,772,057	21,230,743	-2.49
Ophthalmology	339,334	357,273	5.29	10,017,371	10,506,461	4.88
Radiology	623,712	653,850	4.83	9,564,330	9,970,191	4.24
Pathology	2,941,244	3,126,365	6.29	21,369,502	22,489,399	5.24
Obstetrics and Gynaecology	294,288	328,728	11.70	8,774,785	9,151,231	4.29
General practice	4,762,316	4,858,641	2.02	75,806,649	76,545,594	0.97
Totals*	10,938,103	11,351,540	3.78	193,386,798	195,908,988	1.30

** Totals includes some Type of Practice not showned on this table.*

For most categories of medical practice, the estimates drawn from the sample were fairly close to those presented by MH, both for the number of services and for the costs generated in providing these services. The accuracy achieved was all the more remarkable as no post-stratification was carried out at any level to adjust the consumption of health care services to the MH figure.

Table 8. -- Number and Duration of Hospital Stays, 1986-87 Fiscal Year:
Private and Non-Institutional Collective Households

Age Group	Number of Stays			Duration of Stays		
	MH	Sample	Relative Difference (%)	MH	Sample	Relative Difference (%)
0-64	100,127	96,303	-3.82	538,616	499,665	-7.23
65 and over	43,226	41,318	-4.41	452,172	414,555	-8.32
Total	143,353	137,621	-4.00	990,788	914,220	-7.73

Table 8 shows the results of the comparison of the number and duration of hospital stays. Taking the conceptual differences between the two data sources into account, it is deemed to be satisfactory to achieve an accuracy of the estimates within 10%. A larger underestimate for the duration of stays, than for the

number of stays, indicates that longer stays were prone to underestimation. This situation may be explained by the difficulty in identifying residents of institutional collective dwellings on the MH files.

Discussion

With the methodology presented in this article, approximately 74% of the census file corresponding to private households could be matched with the MH file, using mainly age, sex, postal code, family size and family structure. This rate of 74% is satisfactory when compared to the response rate reported in a number of surveys. For example, response rates for the Nova Scotia Nutrition Survey were 79.7% among located respondents and 60.0% for the total sample (Maclean, 1993). The Manitoba Heart Health Survey registered response rates of 77.1% among located respondents and 60.8% for the total sample (Young et al., 1991).

Obviously, considering the various types of errors possible with matching on a large scale, it is not realistic to expect a matching rate of 100%. It is inevitable that the success rate of any probabilistic matching exercise be affected by erroneous data, lags in the collection or updating of the information, as well as conceptual differences between the data sets. Furthermore, while non-matched individuals exhibited different characteristics from matched individuals, rich socio-demographic information concerning the non-matched population was available from the Census. This information was used to select a sample of matches representative of the entire population.

In 95.5% of cases, the pairs formed did associate with the data on an individual's health care utilization and with the socio-economic data on the same individual in the 1986 Census. This rate of accuracy is exceptional, considering that surnames, given names and birth dates were not used in the matching process.

The accuracy obtained in estimating various indicators associated with the consumption of health care (such as mortality, number and costs of medical services, number and duration of hospital stays) justifies the care with which the matching and sampling methods were developed.

In light of the match rates obtained, the rates of concordance of names and the accuracy of the estimates, it can be said that not only is the new database unique in Canada but also that the quality of the data coded in it greatly exceeds that of many surveys based on interviews.

At a time when health expenditures exceed 10% of the GDP in Canada and 13% in the United States, substantial efforts are being made to identify the relationships between health care utilization and health itself. While it is suspected that the level of health perceived by the patient explains a sizable portion of consumption, many studies have focused on consumption by a specific client group, such as the elderly, or on consumption of health care in the years prior to death (Barer et al., 1987; Shapiro and Roos, 1984).

The newly constructed microdatabase opens the door to various studies that were previously not possible in Canada. For example, one of the projects proposed by the MCHPE consists of analysing morbidity with respect to an individual's occupation and by examining the extent to which the health care utilization for a particular class of illnesses is related to the basic occupational group. The census data can be used to classify individuals according to the reported occupation, or according to whether or not they are employed and whether or not they are in the labour force. Using the 9th revision of the International Classification of Diseases (ICD-9), the potential medical conditions to be studied will be musculo-skeletal disorders, cardio-vascular diseases, mental disorders, gastrointestinal illnesses and injuries.

From a general view, there are plans to study differences in the level of health care utilization by socio-economic status at different stages in life. On the one hand, it is well-documented that the greatest consumption of health services occurs toward the end of one's life (Barer et al., 1987; Roos et al., 1987).

On the other hand, a major decline in infant mortality between 1960 and 1990 has also been observed (Pappas et al., 1993; Marmot, 1986). These two phenomena alone are justification for undertaking more thorough comparisons at all age levels. Using data on visits to doctors, health care at home and hospital admissions, it will be possible to compare health care utilization by different age groups by socio-economic status for the classes of illnesses mentioned above. In addition, several studies (Ugnat and Mark, 1987; Williams, 1990; Wilkins et al., 1991) suggest that differences in health conditions according to socio-economic status are greater among persons between 35 and 64 years of age than for other age groups. Analyses by age group using this new database confirmed these hypotheses or shed new light on these matters (Mustard, 1995).

A study to examine the impact of parental socioeconomic status on the use of hospital and ambulatory medical care services during the first year of life has just been completed. It showed that after controlling for low birth weight, maternal age and the joint effects of education and income; for hospital care, education was significantly negatively associated and exhibited a threshold effect between the lowest quartile and all other quartiles; for ambulatory treatment care, income was significantly associated and exhibited a linear effect; for preventive care, both income and education were associated and exhibited a threshold effect between the lowest quartiles and all other quartiles. In the first year of life excluding the birth event, per person public health expenditures were more than twice as high in the lowest education or income quartile compared to the highest quartile (Knighton et al., 1997).

Although the links between socio-economic status and health are the object of intensive research, one of the most frequently encountered problems is that it is impossible to have precise information on socio-economic status at the individual level. Some researchers have no other choice but to use an indicator obtained through the aggregation of taxation or census data for an area of a given size, such as the census enumeration area or the postal code area (Wilkins, 1993). Little research has been done to verify the impact and the validity of this methodology. This tends to reduce the capacity of such models to detect more subtle but theoretically important determinants.

The HALS file, combined with administrative data from health care utilization records, opens the door to comparisons which, until now, have been difficult if not impossible to make. HALS offers us a clear and detailed image of individuals suffering from disability. Whether by age group or sex, by type of disability (mobility, sight, hearing, dexterity, cognition, etc.) or severity, the Manitoba population suffering from a disability can be compared to the general population by means of the census file. Specific analyses of these data focussed on mortality and on health care utilization (Tomiak et al., 1997).

Finally, as the population ages, a greater demand for long-term care services and in particular, nursing homes is expected. A study was initiated to assess the relative importance of predisposing, enabling and need characteristics in predicting nursing home entry.

The list presented here is not exhaustive and is provided only to demonstrate how the new database can be used to analyze health care utilization at different stages in life and for different topics. It is believed that the analytical benefits produced by the record linkage of the administrative sources are important in term of public interest.

Acknowledgments

The authors wish to thank the following persons for their significant and generous contribution to this study: Shelley Derksen, J. Patrick Nicol, and Leonard McWilliam, Manitoba Centre for Health Policy and Evaluation.

Footnote

[1] It should, however, be kept in mind that the match rate on Indian reserves was only 44.5%, considerably lower than the average rate of 74%. This could lead to a bias, since the matched individuals may have had very different characteristics from the reserve population as a whole.

Bibliography

Anderson, G.;Grumbach, K.; Lutt, H.; Roos, L.L.; and Mustard, C. (1993). Use of Coronary Artery Bypass Surgery in the United States and Canada: Influence of Age and Income, *Journal of the American Medical Association*, 269, 1661-1666.

Barer, M. L. et al. (1987). New Evidence on Old Fallacies, *Social Science Medicine* 24(10), 851-862.

David, P. et al. (1993). Linking Survey and Administrative Data to Study the Determinants of Health, *Proceedings of the American Statistical Association*, San Francisco.

Dougherty, G.; Pless, I.B.; and Wilkins, R. (1990). Social Class and the Occurrence of Traffic Injuries and Death in Urban Children, *Canadian Journal of Public Health*, 81, 204-209.

Dolson, D.; Maclean, K.; Morin, J. P.; and Théberge, A. (1987). Sample Design for the Health and Activity Limitation Survey, *Survey Methodology*, 13(1), 93-108.

Health and Welfare Canada (1981). *The Health of Canadians: Report of the Canada Health Survey*, Cat. No. 82-538E.

Fellegi, I.P. and Sunter, A.B. (1969). A Theory for Record Linkage, *Journal of the American Statistical Association*, 64, 1183-1210.

Knighton, T.; Houle C.; Berthelot J. M.; and Mustard C. (1997). The Impact of Socio-Economic Inequity on the Health Care Utilization Practices of Infants During the First Year of Life, Symposium on Intergenerational Equity in Canada, Statistics Canada, Ottawa. (For reprint contact Miles Corak at (613) 951-9047.)

Kovacevic, M. (1995). The Weight Adjustment for the Sample from the "Whole Population Database" (Private Household Component), Technical Note, Statistics Canada, March 24 (not published).

Maclean, D. R. (1993). Report of the Nova Scotia Nutrition Survey, Nova Scotia Heart Health Program, Department of Health, Government of Nova Scotia.

Marmot, M.G. (1986). Social Inequalities in Mortality: The Social Environment. In: *Class and Health, Research and Longitudinal Data*, (Ed. R.G. Wilkinson), London: Tavistock Publications.

Marmot, M.G. and Mcdowall, M.E. (1986). Mortality Decline and Widening Social Inequalities. *Lancet*, I, 274-276.

Mustard, C. (1995). Socioeconomic Gradients in Mortality and the Use of Health Care Services and Different Stages in the Life Course, Research Paper, Manitoba Centre for Health Policy and Evaluation.

Pappas, G. et al. (1993). The Increasing Disparity in Mortality Between Socioeconomic Groups in the United States, 1960 and 1986. *New England Journal of Medicine*, 329, 103-109.

Roos, L.L.; Nicol, J.P.; and Cageorge, S.M. (1987). Using Administrative Data for Longitudinal Research: Comparisons with Primary Data Collection, *Journal of Chronic Diseases*, 40(1), 41-49.

Shapiro, E. and Roos, L.L. (1984). Using Health Care: Rural/Urban Differences Among the Manitoba Elderly, *The Gerontologist*, 24(3), 270-274.

Smith, Martha E. (1981). *Generalized Iterative Record Linkage System*, Health Division, Statistics Canada.

Statistics Canada (1994a). *Health Status of Canadians: Report of the 1991 General Social Survey*, Cat. No. 11-612E, No. 8.

Statistics Canada (1994b). *Causes of Death 1992*, Cat. No. 84-208

Statistics Canada (1988*), The Health and Activity Limitation Survey, Selected Data for Canada, Provinces and Territories*, Cat. No. 41034.

Statistics Canada (1986b). *Report of the Canadian Health and Disability Survey 1983-1984*, Cat. No. 82-555E.

Statistics Canada (1986b). *Census Handbook*, Cat. No. 99-104.

Tomiak, M.; Berthelot, J.M; and Mustard, C. (1997). A Profile of Health Care Utilization of the Disabled Population in Manitoba, (submitted for publication).

Ugnat A M and Mark, E. (1987). Life Expectancy by Sex, Age and Income Level. *Chronic Disease in Canada*.

Young, T.K.; Gelskey, D.E.; Macdonald, S.M.; Hook, E.; and Hamilton, S. (1991). The Manitoba Heart Health Survey: Technical Report.

Williams, D. R. (1990). Socio-Economic Differentials in Health: A Review and Redirection,. *Social Psychology Quarterly* 53(2):81-99.

Wilkins, R.; Adams, O.; and Brancker, A. (1991). Changes in Mortality by Income in Urban Canada from 1971 to 1986, *Health Reports*, 1(2), 137-174.

Wilkins, R. (1993). Use of Postal Codes and Addresses in the Analysis of Health Data, *Health Reports*, 5(2): 157-177, Cat. No. 82-003.

Wolfson, M.C.; Rowe, G.; Gentleman, J.F.; and Tomiak, M. (1993). Career Earnings and Death: A Longitudinal Analysis of Older Canadian Men, *Journal of Gerontology: Social Sciences*.